

Poster III-11

The BioCyc Collection of Pathway/Genome Databases

Karp, Peter D.

Bioinformatics Research Group, SRI International, Menlo Park, CA, USA

A Pathway/Genome database (PGDB) integrates information about the genome, the metabolic network, and the genetic network of an organism. The BioCyc collection of Pathway/Genome DBs available at URL BioCyc.org consists of fifteen PGDBs, including the following.

The **EcoCyc** PGDB, which is a model-organism DB for *E. coli* that tracks the evolving genome annotation of *E. coli*. EcoCyc contains an extensive description of the transcriptional regulatory network of *E. coli*, describing the regulation of 1408 *E. coli* genes in 810 operons, and their control by 125 transcription factors and their corresponding binding sites. EcoCyc also describes the known transporters and metabolic pathways of *E. coli*.

The **MetaCyc** PGDB, which describes 492 experimentally determined metabolic pathways and experimentally studied 1571 enzymes from 194 organisms. MetaCyc is a highly curated reference pathway database for predicting the pathways of other organisms from their complete genomes.

The **HumanCyc** PGDB, which describes 134 predicted metabolic pathways of *H. sapiens*.

The software environment used to implement BioCyc is called Pathway Tools. Pathway Tools is a platform for symbolic systems biology – the development of symbolic reasoning methods for inferring properties of the complete biochemical network of a cell. Among the components of Pathway Tools are: a large biological ontology that facilitates the capture of complex biological knowledge with high fidelity; a set of inference modules for inferring new layers of biological information above an annotated genome, such as for inferring the metabolic pathway map of an organism, and the operons of a bacterium; editing tools that allow scientists to interactively refine and curate these databases; and visualization tools that support web publishing, querying, and analyses of these data. Pathway Tools is in use by more than 30 groups outside SRI who are creating additional PGDBs, such as the AraCyc PGDB for *Arabidopsis thaliana*, available at URL Arabidopsis.org:1555.

Supported by NIH grants GM65466, RR07861, HG02729 and DOE grant DE-FG03-01ER63219 and DARPA contracts N66001-01-C-8011 and F30602-01-C-0153.

References

1. Karp, P.D., Paley, S. and Romero, P., “The Pathway Tools Software,” *Bioinformatics* 18:S225-32 2002.
2. Karp, P.D., Riley, M., Paley, S. and Pellegrini-Toole, A., “The MetaCyc Database,” *Nucleic Acids Research* 30(1):59-61 2002.
3. Karp, P.D., Riley, M., Saier, M., Paulsen, I., Paley, S. and Pellegrini-Toole, A., “The EcoCyc Database,” *Nucleic Acids Research* 30(1):56-58 2002.